Chapter A11: Estimating Benefits with a Random Utility Model

INTRODUCTION

CHAPTER CONTENTS

This chapter describes the random utility model (RUM) and trip frequency model for recreational fishing used in the case study analyses of recreational fishing benefits from the final section 316(b) rule. The model's main assumption is that anglers will get greater satisfaction, and thus greater economic value, from sites where the catch rate is higher, all else being equal. This benefit may occur in two ways: first, an angler may get greater enjoyment from a given fishing trip when catch rates are higher, and thus get a greater value per trip; second, anglers may take more fishing trips when catch rates are higher, resulting in greater overall value for fishing in the region.

EPA relied on two primary data sources in the case study analyses:

- the National Marine Fisheries Service (NMFS) Marine Recreational Fishing Statistics Survey (MRFSS) combined with the Add-on MRFSS Economic Survey (AMES) or combined with the Add-on MRFSS Cost Survey (NMFS 2003a, 2000, and 2003b); and
- the Michigan Recreational Anglers survey, conducted by Michigan Department of Natural Resources (MDNR, 2002).

The North Atlantic and Mid-Atlantic case studies rely on the 1994 MRFSS data; the South Atlantic and Gulf of Mexico case studies rely on the 1997 MRFSS data; and the California case study uses the 2000 MRFSS data. The Great Lakes case study relies on the 2001 MDNR Recreational Anglers survey data. The three datasets provide information on where anglers fish, what fish they catch, and their personal characteristics. When anglers choose among fishing sites they reveal information about their preferences. The case studies use information on recreational anglers' behavior to infer anglers' economic value for the quality of fishing in the case study areas.

EPA used a random utility model to investigate the impact of site characteristics on anglers' site choice for single-day trips. Key determinants of site choice include site-specific travel cost, fishing quality of the site, and additional site attributes such as presence of boat ramps and aesthetic quality of the site. EPA used the 5-year historic catch rates per hour of fishing as a measure of fishing quality in the five coastal region case studies and the Great Lakes regional case study.

The random utility models generate welfare measures resulting from changes in catch rates on a per-trip basis. To capture the effect of changes in catch rates on the number of fishing trips taken per recreational season, EPA combined a RUM model and a trip participation model. The trip participation model estimates the number of trips that an angler will take annually. The combined model is used to estimate the economic value of changes in catch rates or in fish abundance of important fish species in the case study areas.

A11-1 SITE CHOICE MODEL

The site choice model estimates how anglers value access to specific sites, and estimate per-trip economic values for changes in catch rates or fish abundance for different species. The study uses a RUM for its site choice model. The RUM assumes that the cost of travel to a recreational site may be used as a proxy for the "price" of visiting that site. The RUM is therefore a form of travel cost model, using travel costs to estimate economic values for unpriced recreational activities.

The RUM assumes that anglers maximize their utility by choosing the fishing site; mode of fishing (i.e., from shore, private or rental boat, or charter boat); and species that give the greatest level of satisfaction, compared with all available substitutes. Angler *k* chooses site *j* if the utility from that site is greater than the utility from all substitute sites:

$$u_j(k) > u_h(k) \ for \ h \neq j \ and \ h = 1,...J \tag{A11-1}$$

where:

$u_j(k)$ = utility of visiting site *j* for angler *k*;
$u_h(k)$ = utility of visiting a substitute site *h* for angler *k*; and
J = the total number of feasible sites in the angler's choice set.

The RUM travel cost model includes the effects of substitute sites on site values. For any particular site, assuming that it is not totally unique in nature, the availability of substitutes makes the value for that site lower than it would be without available substitutes.

An angler choosing to fish on a particular day chooses a site based on site attributes. The angler weighs the attributes for various "choice set" sites against the travel costs to each site. These travel costs include both the cost of operating a vehicle and the opportunity costs of time spent traveling. The angler then weighs the value given to the site's attributes against the cost of getting to the site when making a site selection.

The RUM therefore assumes that the probability of selecting a particular site is a function of the site attributes, including catch rates, and travel costs to the site:

$$Prob \ (site_j) = f(catch \ rates, \ other \ site \ attributes, \ travel \ cost, \ travel \ time) \tag{A11-2}$$

The RUM assumes that there is a non-random component (v_j) and a random component (ϵ_j) to each angler's utility. The random component is not observable by the researcher (Maddala, 1983; McFadden, 1981). The model therefore assumes that the utility function has a fixed component and a random component, so that:

$$u_j(k) = v_j(k) + \epsilon_j \tag{A11-3}$$

where:

$u_j(k)$ = utility of visiting site *j* for angler *k*;
$v_j(k)$ = the observable component of utility; and
ϵ_j = the random, or unobservable component.

The conditional logit model, most often used to estimate the RUM, is based on the assumption that the random error terms ϵ_j have independently and identically distributed extreme value distributions, and are additive with the observable part of utility (McFadden, 1981; Ben-Akiva and Lerman, 1985).

The logit model therefore becomes:

$$Prob(site_j) = \frac{\exp[v_j(k)]}{\sum_j \exp[v_j(k)]} \ for \ j = 1,\ldots,J \tag{A11-4}$$

where:

$Prob(site_j)$ = the probability that angler *k* will select site *j*;
$\exp[v_j(k)]$ = the angler's utility from visiting site *j*; and
$\Sigma_j \exp[v_j(k)]$ = the sum of the angler's utility for each site, summed over all sites in the opportunity set for a given region.

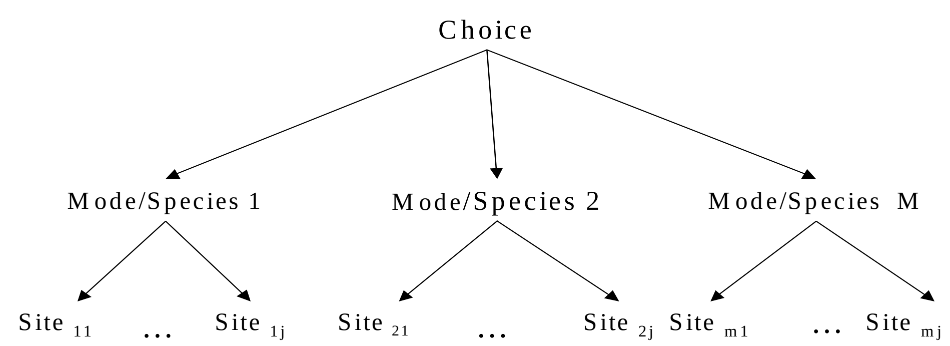
This is estimated as:

$$Prob(j) = \frac{\exp[\beta' x_j]}{\sum_j \exp[\beta' x_j]} \qquad \text{(A11-5)}$$

The conditional logit model imposes the assumption that adding or deleting a site does not affect the probability ratio for choosing any two sites. This so-called independence of irrelevant alternatives (IIA) property follows from the assumption that the error terms are independent (Ben-Akiva and Lerman, 1985). Sites sharing characteristics not included in the model (e.g., saltwater vs freshwater sites) will have correlated error terms, thus violating the IIA property. In these cases a nested logit model, which groups sites with similar characteristics, is more appropriate.

The nested logit model assumes that anglers first choose the group and then a site within that group. Recreational fishing models generally assume that anglers first choose a fishing mode (e.g, fishing from a boat or from shore), and then a site. Thus, the model is structured as a tree, where anglers face a multidimensional choice, consisting of ($M \times J_m$) combinations of modes, *m*, and sites, *j*. The upper levels of the tree — in this case fishing modes — are referred to as branches, while the lower levels — individual sites — are referred to as twigs. This decision tree is illustrated in Figure A11-1.

Figure A11-1: Nested Logit RUM Model



Source: U.S. EPA analysis for this report.

The utility of each element in the angler's choice set is defined as:

$$u_{mj}(k) = V_m(k) + V_j(k) + V_{mj}(k) + \epsilon_m + \epsilon_{mj} \qquad \text{(A11-6)}$$

where:

$u_{mj}(k)$	=	utility of visiting site *j*, fishing by mode *m*, for angler *k*;
$V_m(k)$, $V_j(k)$, $V_{mj}(k)$	=	the observable components of utility; and
ϵ_m, ϵ_{mj}	=	the random, or unobservable components of utility.

In the nested model, ϵ_j is assumed to equal zero, implying there is no correlation across modes, or branches (Ben-Akiva and Lerman, 1985).

The probability that an angler chooses site *j* and mode *m* is:

$$Prob\ (site_j, mode_m) = Prob\ (site_j \mid mode_m)\ x\ Prob\ (mode_m) \qquad \text{(A11-7)}$$

The choice probability for each site is conditional on the choice of mode. Thus, the probability of selecting a site j is

$$Prob(site_j | mode_m) = \frac{\exp[v_{j|m}(k)]}{\sum_{j|m} \exp[v_{j|m}(k)]} \quad j = 1,\ldots,J; \; m = 1,\ldots,M \tag{A11-8}$$

where:

$Prob(site_j \mid mode_m)$	=	the probability that angler k will select site j, given that the angler has selected mode m;
$\exp[v_{j\mid m}(k)]$	=	the angler's utility from visiting site j for mode m;
$\Sigma_{j\mid m} \exp[v_{j\mid m}(k)$	=	the sum of the angler's utility for each site available to mode m, summed over all sites in the opportunity set for a given region and mode;
J	=	the total number of sites available to an angler; and
M	=	the total number of fishing modes (e.g., shore, private boat, or charter) available to an angler.

This is estimated as:

$$Prob(j | m) = \frac{\exp[\beta' x_{j|m}]}{\exp[I_m]} \tag{A11-9}$$

where:

β	=	the matrix of estimated coefficients;
$x_{j\mid m}$	=	the matrix of characteristics of each site j for fishing mode m; and
I_m	=	the expected maximum utility from the choice of a mode, termed the inclusive value for mode choice m.

The inclusive value is defined as:

$$I_m = \log[\sum_{j|m} \exp[\beta' x_{j|m}] \tag{A11-10}$$

The probability of selecting a mode is estimated as:

$$Prob(m) = \frac{\exp[\alpha' y_m + \tau_m I_m]}{\sum_m \exp[\alpha' y_m + \tau_m I_m]} \quad for \;\; m = 1,\ldots,M \tag{A11-11}$$

where:

α	=	the matrix of estimated coefficients on mode characteristics;
y_m	=	the matrix of characteristics of each mode;
τ	=	the coefficient on the inclusive value; and
$\Sigma_m \exp[\alpha' y_{m} + \tau_m I_m]$	=	the sum of the angler's utility for each fishing mode, summed over all modes available to an angler; and
M	=	the number of fishing modes, or branches, in the model.

The coefficient on the inclusive value is related to the correlation between alternatives. The condition $0<\tau<1$ is sufficient for the nested logit model to be consistent with utility maximization (McFadden, 1981). A value of τ between zero and one indicates that there is greater substitutability within, rather than across, groups of alternatives. Thus, there is greater substitutability between sites than across fishing modes. If τ is equal to one, then all modes are equally substitutable, and the model becomes identical to the standard multinomial or conditional logit model, where the IIA property holds for all alternatives.[1]

[1] If consistency with utility maximization is required to hold globally, the coefficients on the inclusive values (i.e., dissimilarity coefficients) must lie inside the unit interval (McFadden, 1981). If consistency with utility maximization is required to hold only locally, then the dissimilarity coefficients can lie outside of the unit interval. In that case, additional tests are required to determine whether conditions for local maximum are satisfied (Kling and Herriges, 1995).

While some of the case study models used the nested logit model, the model was found to be inappropriate for other case studies. In these cases, the conditional logit model was used for site choice estimation. In all of the logit models estimated for the RUM case studies, the measurable component of utility is estimated as:

$$v_j(k) = \beta_1 tc_j(k) + \beta_2 \; tt_j(k) + \beta_3 \; X_j(k) + \sum_s \gamma_s q_{js}(k) \qquad \text{(A11-12)}$$

where:

$v_j(k)$ = the utility realized from a conventional budget constrained utility maximization model conditional on choice *j* by angler *k*;
$tc_j(k)$ = the travel cost to site *j* for angler *k*;
$tt_j(k)$ = the travel time to site *j* for angler *k*;
$X_j(k)$ = a vector of site characteristics for site alternative *j* as perceived by angler *k*. These characteristics may include various site amenities (e.g., presence of boat ramps) and aesthetic quality of the site;
$q_{js}(k)$ = the fishing quality of site *j* for species *s*, measured in terms of catch rate or fish abundance; and
β and γ = the estimated model coefficients.

The study assumes that anglers in the estimated model consider site quality based on the catch rate for their targeted species and additional site attributes, such as the presence of boat ramps or fishing piers. Theoretically, an angler may catch any of the available species at a given site (Morey, 1999). If, however, an angler truly has a species preference, then including the catch variable for all species available at the site would inappropriately attribute utility to the angler for species not pursued (Haab et al., 2000; Hicks et al., 1999; McConnell and Strand, 1994). To avoid this problem, EPA multiplied a dummy variable for each species targeted by the catch rate, so that each angler's observation in the data set includes only the targeted species' catch rate. All other catch rates are set to zero.

A11-2 Trip Frequency Model

The trip frequency model estimates changes in days fished when site or individual characteristics change. The model assumes that the number of days fished in a year is a function of the travel costs, site characteristics, and characteristics of the individual anglers:

$$T = f(p,x,z) \qquad \text{(A11-13)}$$

where:

T = the number of days fished in a year;
p = a vector of travel costs;
x = a vector of site characteristics; and
z = a vector of angler characteristics.

To connect this model to the RUM, the trip frequently model is often specified as:

$$T = f(I(p,x),z) \qquad \text{(A11-14)}$$

where:

I = the inclusive value for each angler, calculated from the RUM;
p = a vector of travel costs;
x = a vector of site characteristics; and
z = a vector of angler characteristics.

The inclusive value can be interpreted as a measure of the expected utility of a set of choice alternatives (Ben-Akiva and Lerman, 1985). The participation model uses the inclusive value from the conditional logit model as a measure of the expected utility of the sites available to anglers in the study region. This is measured by:

$$I_k = \log \sum_j \exp(V_j \ (q_{js})) \tag{A11-15}$$

where:

I_k	=	the inclusive value for fishing sites in the study area for angler k;
$\exp(V_j(q_{js}))$	=	angler's utility from visiting site j; and
q_{js}	=	catch rate for species s at site j.

This study therefore estimates the trip frequency model by first estimating the site choice model (RUM), then using the model results to estimate the inclusive value I_k for each angler. Finally, the study estimates the participation model using the inclusive value and other variables to explain trip frequency. The number of days fished becomes a function of the value per trip, indicated by the inclusive value and individual angler characteristics. This model assumes that changes in site quality and travel costs do not directly influence the number of trips, but that changes in site quality will change trip values, thereby indirectly affecting the number of trips.

The study uses a Poisson regression model to estimate trip frequency. This model is one of those most commonly used for count data: discrete data where the dependent variable is a count or frequency. The Poisson regression model explicitly recognizes the non-negative integer character of the dependent variable (Winkelmann, 2000).

The Poisson regression model assumes the Poisson distribution:

$$f(y_k) = \frac{\exp(-\lambda_k)\lambda_k^{y_k}}{y_k!} \quad for \ y = 0,\ 1,\ 2,\ldots \tag{A11-16}$$

where:

y_k	=	the actual number of trips taken by an individual angler in the sample;
λ	=	both the mean and variance of the distribution (this parameter must be positive); and
k	=	1, 2,...K, the number of individuals in the sample.

If the expected value of the demand for trips in a given time period is E(Y), and:

$$E(Y) = f(I,z,\beta) \tag{A11-17}$$

where:

I	=	the inclusive value;
z	=	a vector of angler characteristics; and
β	=	the vector of estimated coefficients,

then the Poisson probability distribution of demand for trips is:

$$Prob\ (Y_k = y_k) = \frac{e^{-\lambda_k}\lambda_k^{y_k}}{y_k!}, \quad y = 0,\ 1,\ 2,\ldots \tag{A11-18}$$

where:

Y_k	=	the estimated number of trips taken by an individual in the sample;
y_k	=	the actual number of trips taken by an individual in the sample;
k	=	1, 2,...K the number of individuals in the sample; and
$\lambda = f(I, z, \beta)$	=	the expected number of trips for an individual in the sample, where I, z, β are variables affecting the demand for recreational trips (i.e., inclusive value and socio-economic characteristics, and β is the vector of estimated coefficients.

Generally, λ is specified as a log-linear function of the explanatory variables x_i, so that:

$$\ln\lambda_k = \beta x_k \tag{A11-19}$$

or:

$$\lambda_k = \exp(\beta x_k) \tag{A11-20}$$

This function ensures that λ_k will be positive. The parameters of the Poisson regression are estimated by maximum likelihood.

This model's primary limitation is the requirement that the mean equals the variance. The variance often exceeds the mean, resulting in overdispersion. Overdispersion may be viewed as a form of heteroskedasticity (Winkelmann, 2000). If overdispersion exists but the model is otherwise correctly specified, the Poisson estimator will still be consistent. The standard errors will be biased downward, however, leading to inflated t-statistics. When this occurs, researchers often use the negative binomial, which allows for the variance to be greater than the mean. The negative binomial distribution is derived as a compound Poisson distribution, where the Poisson distribution is the limiting form of the negative binomial distribution.

The Poisson model may be modified to derive the negative binomial model by respecifying λ_i so that:

$$\ln\lambda_k = \beta x_k + \epsilon \tag{A11-21}$$

where $\exp(\epsilon)$ has a gamma distribution with mean 1 and variance α (Greene, 1995), yielding the conditional probability distribution:[2]

$$Prob[Y = y_k|\epsilon] = \frac{\exp(-\lambda_k)\ \exp(\epsilon)\ \lambda_k^{y_k}}{y_k!} \tag{A11-22}$$

where:

$Prob[Y = y_k\|\epsilon]$	=	the probability that the estimated number of trips equals the actual number of trips, if ϵ has a gamma distribution with mean 1 and variance α;
y_k	=	0,1,2... number of trips taken by individual k in the sample;
k	=	1,2,..., k number of individuals in the sample; and
λ_k	=	expected number of trips for an individual in the sample.

Integrating out ϵ from equation A11-22 gives the unconditional distribution for y_k, which is used in the model's optimization:

$$Prob\ (Y = y_k) = \frac{\Gamma(\theta + y_k)}{\Gamma(\theta)\ y_k!]\ u_k^{\theta}\ (1 - u_k)^{y_k}} \tag{A11-23}$$

where:

$Prob(Y = y_k)$	=	the probability that the estimated number of trips equals the actual number of trips;
y_k	=	0,1,2... number of trips taken by individual k in the sample;

[2] EPA chose this particular parameterization because it is used by the LIMDEP™ software package.

$\Gamma(.)$ = gamma function;[3]
θ = $1/\alpha$, where α is an overdispersion parameter; and
u_i = $\theta/(\theta + \lambda)$.

The negative binomial model has an additional parameter, α, which is an overdispersion parameter, such that:

$$Var\ [y_k] = E\ [y_k]\ (1 + \alpha\ E\ [y_k]) \quad \text{(A11-24)}$$

The overdispersion rate is then given by the following equation:

$$Var\ \frac{y_k}{E\ [y_k]} = 1 + \alpha\ E\ [y_k] \quad \text{(A11-25)}$$

(Greene, 1995).

EPA used the negative binomial model to predict the seasonal number of recreation trips for each recreation activity based on the inclusive value, individual socio-economic characteristics, and the overdispersion parameter, α. If the inclusive value (i.e., the measure of the expected utility of site alternatives) has the anticipated positive sign, then increases in the inclusive value stemming from improved fishing quality at the sites in the study area will lead to an increase in the number of trips. The combined multinomial logit (MNL) model site choice and count data trip participation models allowed the Agency to account for changes in per-trip welfare values, and for increased trip participation in response to improved ambient water quality at recreation sites.

A11-3 WELFARE ESTIMATION

The case studies estimate changes in economic values when catch rates for different species change. Changes in catch rates will affect economic values in two ways. First, the value per trip will change; and second, the number of trips taken may change. The study measures the total economic value for a change in the quantity or quality of particular sites by the number of days fished per angler times the economic value per trip per angler. This value varies with the quality and number of available sites. The total value of a change in catch rate is measured as:

$$TEV = N \times X \times WTP \quad \text{(A11-26)}$$

where:

TEV = the total economic value for a specified period of time, such as a season or year;
N = the number of participants;
X = the number of trips per participant; and
WTP = the value per angler per trip, measured by the amount of money that the angler would be willing to pay for a fishing trip.[4]

The study first estimates the value per trip using the RUM, and then estimates the number of trips per angler using the trip frequency model. The results of these models must be combined to measure the total economic value for a given change.

The value of an improvement in site quality, in this case the catch rate or fish abundance, can be measured by the compensating variation (*CV*) that equates the expected value of realized utility under the baseline and post-compliance conditions. If the catch rate increases from q^0 to q^1, then the *CV* will be measured by:

$$v_j(p_j,\ q_j^1,\ y - CV) + \epsilon_j = v_j(p_j,\ q_j^0,\ y) + \epsilon_j \quad \text{(A11-27)}$$

[3] Gamma function is a notation for a definite integral that appears in the equation. For detail on gamma function see Mood et al. (1974).

[4] The estimated model and resulting welfare estimates rely on the assumptions that the number of participants is fixed in the short run, and that the value per trip is independent of the number of trips.

where:

p_j = the fishing price, or travel cost, for site *j*;
q_j^1 = the quality, measured by catch rate, for site *j* under the post-compliance conditions;
q_j^0 = the quality, measured by catch rate, for site *j* under the baseline conditions; and
y = the angler's income.

To calculate *CV*, the angler's utility $(V_j(k))$ must be estimated as a function of price, quality, and income. The marginal utility of income cannot be estimated in the logit model because each angler's income does not change across alternatives. Price (travel cost), however, enters the indirect utility function *V(j)*, so that the model can assume the estimated coefficient on travel cost to be the negative of the marginal utility of income (Bockstael et al., 1991).

The RUM predicts only the probability of choosing a specific site. The measure of *CV* must therefore account for the researcher's uncertainty in predicting site choice. Measuring *CV* in terms of expected value yields:

$$E[v(p,\ q^1,\ y - CV)] = E[v(p,\ q^0,\ y)] \tag{A11-28}$$

where:

$v(p, q, y)$ = expected maximum utility of being able to choose among *J* sites on a given fishing trip;
p = the fishing price, or travel cost;
q^1 = sites' quality, measured by catch rate, under the post-compliance conditions;
q^0 = sites' quality, measured by catch rate, under the baseline conditions; and
y = the angler's income.

If the marginal utility of income is assumed to be constant, the compensating variation for the logit model is (Bockstael et al., 1991; Parsons et al., 1999):

$$\begin{aligned} CV_k &= (-1/\beta_1)\ [[\ln\sum \exp[v_j(q^1)] - \ln\sum \exp[v_j\ (q^0)]] \\ &= (-1/\beta_1)\ [I^1 - I^0] \end{aligned} \tag{A11-29}$$

where:

CV_k = the compensating variation for individual *k* at site *j* on a given day;
j = 1,...,J represents a set of alternative sites in the study region;
β_1 = the negative of the marginal utility of income, measured by the coefficient on travel cost;
I^0 = the baseline inclusive value; and
I^1 = the post-compliance inclusive value.

CV for the nested logit model is calculated as follows (Bockstael et al., 1991; Hicks et al., 1999):

$$\begin{aligned} CV_k &= \frac{\ln\left[\sum_m \left(\sum_j \exp\left(\dfrac{V_{jm}(q^1)}{(1-\tau)}\right)\right)^{(1-\tau)}\right] - \ln\left[\sum_m \left(\sum_j \exp\left(\dfrac{V_{jm}(q^0)}{(1-\tau)}\right)\right)^{(1-\tau)}\right]}{-\ \beta_1} \\ &= (-1/\beta_1)\ [I^1 - I^0] \end{aligned} \tag{A11-30}$$

where:

CV_k = the compensating variation for individual k at site j on a given day;
j = 1,...J represents a set of alternative sites in the study region for fishing mode m;
m = 1,...M represents alternative fishing modes available to an angler;
β_1 = the negative of the marginal utility of income, measured by the coefficient on travel cost;
I^0 = the baseline inclusive value; and
I^1 = the post-compliance inclusive value.

This gives the expected compensating variation for a choice occasion. To obtain the value per season, EPA multiplied the *CV* per trip by the number of trips estimated with the participation model. The two models are linked through the inclusive value, which weights the indirect utilities associated with different sites and their prices and qualities by the probabilities of choosing each site (Bockstael et al., 1991).

Parsons et al. (1999) compared several models that link site choice and trip frequency models, and find that they produce similar welfare estimates. Two methods for estimating seasonal welfare estimates are relevant to the models estimated in this case study. The first, proposed by Bockstael et al. (1987), calculates the per-trip welfare measure from the RUM, using the measure of *CV* presented above (Eq. A11-30). The authors then use the trip frequency model to predict the change in the number of trips taken under the proposed policy change. Finally, they calculate a seasonal welfare measure in one of two ways:

$$W_{low} = CV \times Pred(T^0) \tag{A11-31}$$

$$W_{high} = CV \times Pred(T^1) \tag{A11-32}$$

or

$$W = CV \times \frac{[Pred(T^0) - Pred(T^1)]}{2} \tag{A11-33}$$

where:

W_{low} = the low bound estimate of the seasonal welfare gain;
W_{high} = the upper bound estimate of the seasonal welfare gain;
CV = the compensating variation for an individual on a given day;
$Pred(T^0)$ = the predicted number of trips before the policy change; and
$Pred(T^1)$ = the predicted number of trips after the policy change.

The second method, based on Hausman et al. (1995), calculates seasonal welfare based on the trip frequency model.

EPA used the first method (Bockstael et al., 1987) to estimate lower and upper bound values for the seasonal welfare gain per individual. The Agency extrapolated the estimates of seasonal value per individual to the regional level based on estimates of the total participation level in the region. Procedures for estimating total regional participation are case study specific and discussed in the relevant chapters.

A11-4 DATA SOURCES

The data used for the regional case studies of recreational benefits are from the NMFS MRFSS in the Southeastern, Northeastern, and California regions in the U.S., and the MDNR Recreational Anglers survey database. The following sections provide a general description of each data source, sampling methods, and key variables. More detailed information on the sub-sample used in each case study can be found in the relevant case study sections.

A11-4.1 Marine Recreational Fisheries Statistics Survey

MRFSS is a long-term monitoring program that provides estimates of effort, participation, and finfish catch by recreational anglers. The MRFSS survey consists of two independent, but complementary, surveys: a random digit-dial telephone survey of households and an intercept survey of anglers at fishing access sites. Sampling is stratified by state; fishing mode (shore, private/rental boat, party/charter boat); and wave, and allocated according to fishing pressure. Fishing sites are randomly selected from an updated list of access sites.

The intercept survey distinguishes between the modes of fishing (i.e., shore, private/rental boat, party/charter boat), and is designed to elicit information about fishing trips just completed by anglers. The basic intercept survey collects information about anglers' home zip code, the length of their fishing trip, the species they were targeting on that trip, and the number of times anglers have been fishing in the past 2 and 12 months. Trained interviewers record the species and number of fish caught that are available for inspection, and weigh and measure the fish. Anglers report the number and species of each fish they caught on the trip that are not available for inspection (e.g., fish that were released alive or used for bait). The intercept survey provides the information used to estimate the historic catch rates at the case study sites for the individual species.

The random telephone survey is used to estimate the number of recreational fishing trips during a 2-month period (as opposed to annual participation) for coastal households. Households with individuals who have fished within 2 months of the phone call are asked about the mode of fishing, the gear used, and the type of waterbody where the trip took place for every trip taken within that period. NMFS estimates total catch and participation by state using the MRFSS telephone and intercept surveys, combined with U.S. Census Bureau and historical data (NMFS, 2003b). The effort estimates (i.e., number of trips) are used in the economic valuation work to expand mean trip-level recreational fishing values to aggregate, population values for recreational fishing. More details about the intercept and the random phone surveys can be found in the MRFSS Procedures Manual (NMFS, 1999a).

NMFS supplemented the routine MRFSS with socio-economic data from anglers in Southeastern and Northeastern regions.[5] The economic survey (AMES) was designed as an add-on to the MRFSS to take advantage of sampling, survey design, and quality control procedures already in place. Economic questions were added to the intercept survey, and a follow-up survey conducted over the telephone was designed to elicit additional socio-economic information from anglers who completed the add-on economic intercept survey. The AMES was implemented from Maine to Virginia in 1994 and from North Carolina to Louisiana in 1997.

The economic field intercept survey of anglers solicited data about trip duration, travel costs, distance traveled, and on-site expenditures associated with the intercepted trip. The survey was conducted by a private survey firm and administered to all marine recreational anglers aged 16 and older intercepted in the field. Data were collected according to the field sampling procedures specified in the MRFSS Procedures Manual. The economic questionnaire was administered either at the completion of the routine MRFSS questions (before inspection of fish) or after all available fish were identified and biological measurements had been obtained. As in the MRFSS, all survey participants, with the exception of beach-bank shore anglers, must have completed their fishing for the day.

Anglers were screened for willingness to participate in the telephone follow-up survey at the time of field intercept. Only those anglers agreeing to the add-on economics field survey or a telephone follow-up survey were interviewed. The telephone follow-up survey solicited additional data and information about anglers' recreational fishing avidity, attitudes, and experience.

A total of 14,868 follow-up surveys were attempted in the Northeast region in 1994, of which 8,226 (55 percent) were completed. Refusals, wrong numbers, and households that could not be reached in four calls accounted for the 45 percent non-response rate. The 1994 questionnaire targeted two distinct groups of anglers: (1) anglers who targeted — not merely caught — bluefish, striped bass, black sea bass, summer flounder, Atlantic cod, tautog, scup, or weakfish; and (2) anglers that targeted other species and happened to catch any of these 8 species. These species were chosen because they were either under management in 1994 or were expected to come under management in the near future. Approximately 10,000 AMES telephone interviews were completed in the Southeast region in 1997. The interview consisted of anglers intercepted from March 1997 through December 1997 and who agreed to be interviewed. More extensive details regarding the final results of the telephone follow-up survey are provided in Hicks et al. (1999).

[5] Socio-economic data are not available for the California region.

The Agency used data from the 1994 and 1997 AMES to model recreational fishing behavior in the Northeastern (including North Atlantic and Mid-Atlantic) and Southeastern (including South Atlantic and Gulf of Mexico) regional case studies, respectively.

A11-4.2 Michigan Anglers Survey

The Great Lakes regional case study used data from the 2001 MDNR Recreational Anglers survey (Lockwood et al., 1999). The MDNR Fisheries Division uses roving and access site angler survey methods to collect angling effort and catch or harvest information from Inland and Great Lakes fisheries. These surveys follow a stratified design using structured sampling within strata. The collected data reflect angling characteristics for specific locations during specific calendar and daily periods. The Michigan angler surveys consist of two separate sampling components: interviews of angling trips and counts of anglers. Interviews collect information on the number of anglers in the party, length of the fishing trip, targeted species, catch or harvest by species, site information, angling mode, and zip code of angler's home town. Typically, angling information is collected by individual angler for roving surveys, not by angling party, to avoid angler party size bias. Angler surveys provide information on both the number of fish harvested and the number of fish caught and released. However, because for caught-and-released fish, neither species type nor number are observed, the estimates of caught-and-released fish are subject to recall, prestige, and rounding errors, and species of fish are more likely to be misidentified. Angling effort is reported as estimated angler-hours or estimated angler-trips. Angler-hours reflect total hours from arrival at a given site to departure from that site for a given time period. Angler trips correspond to the number of times anglers fish at a given location for a given period of time. Part G, Chapter G4 of the Regional Studies Document provides descriptive statistics for the Michigan Angler survey.

A11-5 Limitations and Uncertainties

Recreational survey results may suffer from recall bias, non-response bias, and bias due to sampling effects:

- Recall bias can occur when respondents are asked the number of days in which they recreate over the previous season, such as in the NDS survey. Some researchers believe that recall bias tends to lead to an overstatement of the number of recreation days, particularly for more avid participants. Avid participants tend to overstate the number of recreation days, since they count days in a "typical" week and then multiply them by the number of weeks in the recreation season. They often neglect to consider days missed due to bad weather, illness, travel, or when fulfilling "atypical" obligations. Some studies also found that the more salient the activity, the more "optimistic" the respondent tends to be in estimating the number of recreation days. Individuals also have a tendency to overstate the number of days they participate in activities that they enjoy and value. Taken together, these sources of recall bias may result in an overstatement of the actual number of recreation days.

- Non-response bias. A problem with sampling bias may arise when extrapolating sample means to population means. This could happen, for example, when avid recreation participants are more likely to respond to a survey than those who are not interested in the forms of recreation, are unable to participate, assume that the survey is not meant for them, or consider the survey not worth their time.

- Sampling effects. Recreational demand studies frequently face two types of observations that do not fit general recreation patterns: non-participants and avid participants. Non-participants are those individuals who would not participate in the recreation activity under any conditions. Assuming that an individual is a non-participant in a particular activity if he or she did not participate in that activity at any site tends to understate benefits, since some individuals may not have participated during the sampling period simply by chance, or because price/quality conditions were unfavorable during the sampling period. Avid participants can also be problematic because they claim to participate in an activity an inordinate number of times. This reported level of activity is sometimes correct, but often overstated, perhaps due to recall bias. These observations tend to be overly influential in the model and may lead to overestimation of the total number of trips.

The RUM analyses rely on the unweighted MRFSS data, not correcting for stratification. The MRFSS data is prone to avidity bias where the probability of being interviewed increases with the number of fishing trips (Thomson, 1991). EPA did not correct for avidity bias, which may result in overestimation of the predicted number of trips per season. This bias is unlikely to have a significant effect on benefit estimates, because the predicted number of trips was used only for estimating changes in fishing participation due to improved fishing opportunities. The estimated change in the number of trips was very small (see

Chapters C4, D4, E4, and F4 of this report for detail). The baseline level of participation used in the analysis was taken from NMFS. This estimate was corrected for avidity bias by NMFS.

Similarly to the MRFSS data, the Michigan Angler Survey data are prone to avidity bias where the probability of being interviewed increases with the number of fishing trips (Thomson, 1991). This may result in overestimation of the reported number of trips per season. In addition, the estimates of caught-and-released fish are subject to recall, prestige, and rounding errors, and species of fish are more likely to be misidentified. The effect of this bias on benefits estimates is, however, uncertain.